DOCUMENT RESUME

ED 111 155 95 EC 073 549

AUTHOR Turner, Richard L.
TITLE Research Bases for Performance-Based Teacher Education Options: Supplement to Design for Competence Based Education in Special Education.
INSTITUTION Syracuse Univ., N.Y. Div. of Special Education and Rehabilitation.
SPONS AGENCY New York State Education Dept., Albany. Div. for Handicapped Children.; Office of Education (DHEW), Washington, D.C.
PUB DATE May 74
GRANT OEG-0-73-6140
NOTE 20p.; Paper presented at the CBTE Conference (Glenmont, New York, May 14-16, 1974). For related information, see ED 107 049

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS Conference Reports; *Curriculum Design; Effective Teaching; Exceptional Child Education; *Handicapped Children; *Performance Based Teacher Education; Program Planning; *Research Needs; *Research Reviews (Publications); Special Education; Teacher Certification

ABSTRACT

Intended to supplement previously reprted conference proceedings on the preparation of special educators, the paper presents an overview of the structure of performance-based teacher education (PBTE) and certification, reviews research on selecting relevant program options, and identifies future research needs. Briefly covered are the following structural phases: selection, training, certification, and process and product criteria for evaluating work success. The author points to limited research bases for selecting one PBTE program option over another and recommends using the "General Catalog of Teaching Skills" as one approach for identifying skills and knowledge that should be incorporated into teacher education programs. (LH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED111155

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Supplement

to

Design for Competence Based Education in Special Education

Richard L. Turner, Professor of Education and Associate Dean for Research and Development at Indiana University, Bloomington, presented the paper, *Research Bases for Performance-Based Teacher Education Options* at the CUNY-UFT Conference on CBTE at Tarrytown on March 2, 1974. The United Federation of Teachers and the City University of New York gave permission to reprint this copyrighted paper which represents substantially the material Dr. Turner presented to the Glenmont conference participants on May 15, 1974. The editors of *Design for Competence Based Education in Special Education* are pleased to present this reprint as an integral attachment to the Proceedings from the New York State Conference on Competence Based Teacher Education and Special Education.

This reprint is a product of Grant No. OEG-0-73-6140 funded by the United States Office of Education to the Board of Regents of the University of the State of New York and administered by the State Education Department, Division for Handicapped Children.

The CBTE Conference was conducted at Glenmont, New York on May 14-16, 1974 in cooperation with the Division of Special Education and Rehabilitation of Syracuse University.

EC073549

Richard L. Turner

RESEARCH BASES FOR PERFORMANCE-BASED TEACHER EDUCATION OPTIONS

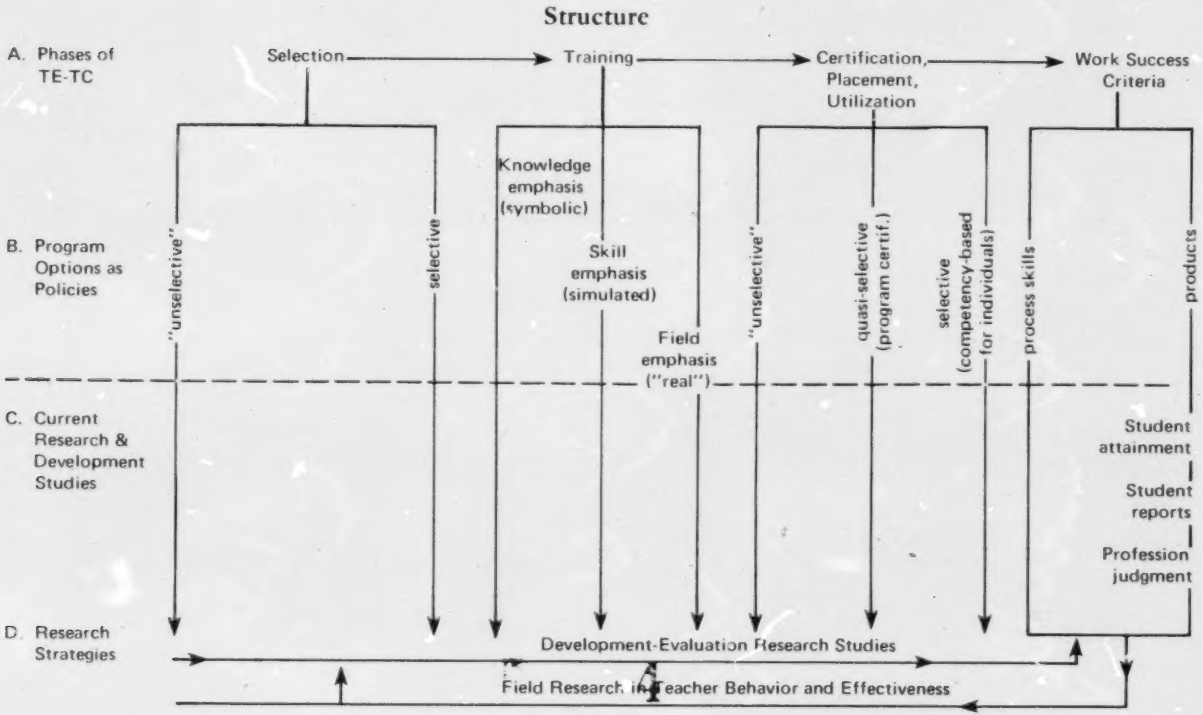
In talking with you this afternoon I will be addressing two related topics. The first is an overview of the structure of teacher education, followed by a review of some of the research germane to various points in this structure. The outcome of this topic should be a clearer idea of the various options available in performance-based teacher education, and a notion of how research findings bear on these options. The second topic picks up on a question left unresolved in the first topic – finding a rational way to select among all of the skills teachers might need to function effectively in a highly pluralistic society, in which various parent groups and their attendant children, various schools, and perhaps, various power blocks, wish to see evidence of prescribed types of teacher behavior. In addressing the latter topic, I will refer principally to the approach we took to this question in constructing the *General Catalog of Teaching Skills* which we recently completed at Indiana University.

Overview

The "structure" of teacher education or a paradigm of it, which I will just outline now and elaborate on later, will serve as a guide throughout the discussion. (Refer to the figure on page 168.) Part of this structure deals with the four phases of teacher education. In Selection, the first phase, there is a choice-point – here one can choose to be "unselective," within some limits, or selective. From this point, one usually enters a training phase for which there are three options. Most people experience "knowledge training" – that is, they engage in a symbolic or verbal activity – in their teacher education programs. They may also engage in skill development which in our program and the Stanford University system is frequently referred to as "simulation." During this process, the trainee builds up classroom skills to a functional but not polished level. In field experience, the third option, the trainee integrates knowledge and skills so that his behavior is more polished, albeit not perfect.

Richard Turner
Indiana University
3/2/74 NYUFT

RESEARCH BASES FOR PBTE-TC OPTIONS



After the training – or in the latter part of the training, certification becomes another choice-point, along with the placement of the teacher. This choice is of initial importance, especially to teachers, because of the different teaching contexts in which they may find themselves in a pluralistic society. The schools they enter, whether urban or rural, inner city or suburban, working class or middle class, will influence the effectiveness of their skills. McDonald has taken this matter into account to a considerable degree in his study.

(Editors' Note: Dr. Turner here makes reference to a presentation at the CUNY-UFT Conference by Dr. Frederick McDonald.)

"Utilization," a term now used at the National Institute of Education, is conjoined to placement, but focuses specifically on how the talents are used in a particular position. In Indiana, it is not uncommon for a teacher to have a major certification field and yet teach in a field in which he or she is minimally qualified. For example, an athletics coach might be teaching history or mathematics. A major issue in certification, then, is the job you are certifying someone to do. Will a teacher who has been certified primarily to coach be competent to do everything else that coaches might be asked to do in a small school?

The final phase deals with work success criteria. We can look at that in several different ways. When McDonald talked about teacher performance, he meant process criteria – acts which the teacher performs, the interactions of teachers and pupils. He also discussed student performance, which meant student achievement, as one kind of product criterion for teacher performance. We have done some studies on other work success criteria, one of which relates to student appraisals or reports of teachers – that is, student evaluation of instruction. I consider this a product criterion because what students think of their instruction is an outcome of that instruction.

A third product criterion is "professional judgment" or the professional reputation of a teacher among the administrative personnel, the teacher's peers, a college supervisor, or anyone who can render a professional judgment about teaching. The reputation that a teacher holds, in other words, is another product of his or her teaching and is not just a process outcome, as many people believe. This reputation may result in part from the duties teachers perform in the school other than classroom teaching.

Under Program Options as Policies (B), one has a choice of placing different kinds of emphases on the options. For example, "non-selection" could be emphasized under Selection, simulated skills under Training, and so forth. Later I will discuss some of the research data that relates to each of these choices together with the different ways we can do research studies bearing on the choices.

At this point I might note that there are two basic research strategies. The first one, Development-Evaluation Research Studies (D), may be most easily described here as making predictions about program outcomes. Let us suppose that we are running two programs. In one, we select the students on the basis of some criterion, and in the other, we have not selected the students. After the completion of these

two programs, we might study the immediate outcomes, but implicitly we must make a prediction about the success of the unselected versus the selected people in their later work.

In another study, we might introduce a heavy knowledge emphasis into a program, together with a low skill emphasis, and look at the outcomes of that program versus one which reverses the emphasis. In both cases, the researchers would be developing the treatments first in order to find out the consequences. This type of research involves follow-up studies and has some of the qualities of experimental field studies. In both studies mentioned, the researcher must implicitly predict that making a change in current practice will have good outcomes.

The second strategy, Field Research in Teacher Behavior and Effectiveness, which McDonald illustrated in his discussion, primarily involves the relationships between process criteria and product criteria. These studies – and there have been a fair number of them – make the assumption that if the skills now seen in operation were to be taught to the teacher in pre-service training, these skills would be effectively implemented at a later time. Let us suppose, for example, that we have assessed the behavior or skills of many third grade teachers and find positive correlations between certain skills and selected pupil outcomes. From these results, for purposes of teacher education, we must now assume that if teachers were trained in these skills at the pre-service level the skills-as-trained would be correctly exercised later and that positive pupil outcomes would follow. This assumption, however, is rarely tested.

Thus, there is a gap in this strategy, and there is not much evidence that we can successfully jump over it. By measuring aptitudes, the McDonald study is building in some protection, because aptitudes are more stable than specific behaviors or skills. The inductive and deductive abilities and vocabularies of individuals change less than their day-to-day or year-to-year classroom behaviors. Nonetheless, the study must make an inferential leap from measures of teacher behavior in a classroom at any one point back to the proposition that those behaviors are trainable and will later be used in the same form in which they were trained.

Selection

Now that we have an overview of the structure, I want to return and examine some of the research bases connected to the two options under Selection. We can cast the first option in question form: If admission into a program is unselective, how will things tend to turn out? One outcome is that there will be more girls in the program than boys. In addition, the girls will show more commitment to teaching than the boys will. Girls also tend to choose at an earlier age to teach in elementary school (Balzer, 1966). They apparently have decided that this role fits them and have carried that idea or goal through into a teacher education program. This development, by the way, is an interesting aspect of sex role acquisition, which we should study in greater depth.

A second outcome shows that girls who enter secondary school teaching tend to make this choice during their high school years. Boys generally do not choose this

course at all. They tend to delay their commitment to teaching a long time, and as some studies have shown, frequently until placement time (Perry, 1972). To males, teaching often represents an alternative which they see as perhaps leading to other kinds of positions or jobs at a later time. Teaching is a career which boys frequently believe they should choose not to enter if they can avoid it, or can opt out of if they want to. Thus, present studies suggest that there is much less commitment to the profession among males than there is among females.

A study conducted at Louisiana State University (Ducharme, 1970) revealed positive outcomes for teachers who had had voluntary experiences working with children before they entered their preparatory program. Administrative officers in the schools rated these teachers higher than those without such experience, thus demonstrating a positive correlation between early teacher work experiences and subsequent administrative ratings. Secondly, these teachers tended to get somewhat better achievement gains from their students than those not having such experience. These findings probably relate to early interests of the teacher, who has been practicing some form of the teaching role before entering a preparatory program. We should look at these outcomes more closely. I am not sure that *asking* someone to get this early experience would produce the same effect. If a program *requires* early work experience with children, then everybody will have it whether they like it or not, and no correlation will exist.

A researcher at the United States International University (Baker, 1970) also conducted a study bearing on Selection. In it students were given intake interviews during which they were rated for potential success. After these students entered the field as teachers, the researcher asked their supervisors or principals to rate them. The outcomes indicated a correlation between the intake interview ratings of the students as freshmen and the subsequent ratings of them as teachers by the administrative officials. This leads me to believe that there may be some consistency in professional judgment. That is, if a committee of professionals handled the selection operation for a teacher training program, they might be able to pick people who would subsequently be seen as successful teachers by other professionals.

A study conducted in Brooklyn by Shirley Weeden (1972) beginning in the mid-1960's showed, however, that out of the hundred or so students dismissed from the Brooklyn preparatory program over a period of four or five years, about eighty percent found their way into teaching one way or another. It seems clear from this study that one cannot easily stop up the holes in the teacher training system. If New York State programs eliminated all of the trainees they believed would be unsuccessful, these trainees would probably go to some other state and get their teaching certificates there. Eliminating students at certification time may or may not be more effective. Again, the students may come in by the back door, and the state will again be back to the certification problem it thinks it has solved.

A number of currently used measures are not very good predictors of a person's work success. If given early, achievement tests, grade point averages, and other intellective measures usually do not predict work success (Ducharme, 1970, Baker,

1970), although there is some contrary evidence (Balzer, 1966). In the Balzer study data, girls who had decided early to teach school and had entered teachers colleges turned out to be superior, on seven of the eight variables measured, to girls who had decided much later to teach. These variables were predominantly intellective items such as vocabulary, language ability, and mathematics. However, the study did not follow these girls far enough to tell us whether or not they made better teachers. Although we all engage in some kind of screening-out procedures – for example, by not admitting students with below C averages into the program, we do not get very good predictions with the ones accepted above that level. McDonald, however, may get better results in his study because his measures are more specific than grade point averages or general achievement tests.

One variable we looked at at Indiana seems promising (Yoder, 1967, Turner, 1967). It is called the "Organized, Business-Like Behavior" versus "Slipshod, Irresponsible Behavior," and appears in scales originally developed by D. G. Ryans. We tested trainees on these scales prior to student teaching and then looked at those who had discipline problems during their field experience.

The results of this study indicated that the student teachers who scored high in organized, business-like behavior had significantly fewer discipline problems than those who scored low, i.e. were slipshod and less responsible. We also found the same results in a predictive study of first-year teachers with the more business-like teachers having substantially fewer discipline problems during the first year.

It would be possible, I think, to generate some criteria by which selection into a program can be controlled. Financially speaking, there are sound reasons to do so. Let us take, for example, a student who has gone through three years of a program, but is eliminated in the fourth. Whether we like it or not, we have run up that student's costs and probably affected the rate of return on his education. It is certainly more humane to eliminate people relatively early in the program than to allow them to continue and then eliminate them at the end. In any event, I hope that in performance based teacher education we will think about the possibility of selecting potentially successful people rather than placing all of the emphasis on training.

Training

Training, as an activity of Development and Evaluation Research Studies, is a process that brings about an outcome. In a program which emphasizes knowledge, we need to know whether the level of knowledge a student attains as an outcome of his program is in fact a good predictor of teacher success. Neither college or high school grade point averages are good predictors. One study (Baker, 1970) found that although college GPA's correlated with student teaching grades, they did not correlate with work success ratings by professional judges. Standardized achievement tests given to teachers seem to be non-predictors of the attainments of teachers. Paper and pencil tests, such as final examinations in methods courses, also usually do not predict teacher success.

Problem-solving tests used to measure teacher knowledge do show some relationships to outcomes, but they have to be done in a particular way. A performance test which I helped develop in the early 60's contained one section which presented examples of pupil work from which teachers had to diagnose pupil difficulties. In another section, we asked teachers to organize a group of elementary arithmetic problems in the order of their difficulty for the pupils. Among the various test items we used, the teacher's ability to diagnose student difficulties and to organize materials turned out to be one of the best correlates of arithmetic gain in the classroom (Turner, 1964). Sequencing ability, or an ability to organize material hierarchically also turns out to be an aspect of knowledge training which is a reasonable predictor of teaching success (Trojcak, 1969). Therefore, the point here is not whether teacher knowledge correlates with student learning, but *how* we get at the knowledge of the teacher as a predictor of success. A teacher's ability to use knowledge to diagnose pupil problems and to sequence materials seems to correlate with student outcomes.

The training and protocol movement offers a third way for developing knowledge. Instructors within this movement try to teach a concept to trainees by showing its instances in the classroom, and then they evaluate whether the trainee has learned the concept in an applicable way. For example, a trainee would learn to recognize instances of a particular kind of behavior from a film showing real classroom situations. Such a film might depict examples of positive or negative reinforcement, shaping, or any of a wide variety of concepts. Protocol materials, then, stress that the trainee apply rather than merely verbalize knowledge.*

Trainees, as they viewed this film would code approval or disapproval and praise or rejection behaviors. This kind of training treatment provides a better test of the knowledge of concepts used in teaching than most of the tests we now have.

We followed the use of protocol films with training materials which asked the teacher to produce the behavior that the concept symbolized. That is, the teacher would actually have to demonstrate behaviors depicting approval, praise, reinforcement, criticism, and rejection. Through this method we can test the level of skill a trainee has before he or she goes into classroom situations. It is accurate to say that we can increase that person's skill level prior to classroom performance, but final polishing of the skill requires classroom practice.

In a research seminar I conduct on the assessment of teacher performance, one of my students raised the question, "What are the characteristics of approval behavior, and how can we array those characteristics on another rating scale to show the *quality* of approval or disapproval behavior?" This student will spend about a year working on this problem. Finding solutions to such questions requires tremendous amounts of time and money. The McDonald study is trying to find out whether such actions as approval or disapproval make any difference. To do so,

*At this point in the presentation, Dr. Turner ran a film clip showing brief instances of classroom situations in which teachers demonstrated the behaviors or concepts of approval and disapproval, praise and rejection. He explained beforehand, however, that these concepts had not been proven to be highly effective – that is, "approval" on the one hand or "disapproval" on the other did not invariably correlate with pupil outcomes.

these researchers will have to develop a special scale which includes all those attributes showing the quality of approval – disapproval and correlate that to pupil attitudes, achievement, and other outcomes. At the present time, Performance-Based Teacher Education forces us to *assume* the answer to such questions, and I do not think we can now assert what the final research outcome will be.

Experimental studies (Hiller, 1972) have shown that teacher vagueness is a characteristic which correlates, in the knowledge domain, to pupil outcomes. Vagueness, for example, can be controlled in a speaker by depriving the speaker of information. If one provides trainees with only a small amount of information on a subject and asks them to give a speech on that topic, they will, of necessity, have to use a pretty vague approach simply because they do not know very much. But if one gives them more information, their speech will be less vague. Vagueness, then, is partly a function of the amount of knowledge a person has. I do not want to become excessively political here, but vagueness ratings of the Nixon-Kennedy debate showed that Nixon was more vague than Kennedy. Without carrying the matter too far, I will just add that it is of course possible to be vague for reasons other than a lack of knowledge.

McDonald, together with others, has created at Stanford University one of the very good skill-development systems. Although he is more expert in this area than I am, I will take the liberty to briefly discuss some of the things he has done. To use his expression, this explanation is accurate but not necessarily precise. In this system, the trainee has a video tape made of his or her own performance in a microteaching situation. The instructor views this tape to see what the teacher, without training, can do. The instructor then shows the trainee a model of a teacher engaging in a specific kind of behavior that the trainee should be able to perform, such as praise or approval, or asking higher order questions. The trainee analyzes and identifies the components of that model teacher's behavior and compares it to his or her own videotaped performance. Then the trainee teaches again and the istructor observes the new performance on videotape and appraises the extent to which the trainee has improved in the skill since the pre-test entry point, or the first videotape, as the result of the interim training. In this case, the instructor uses a modeling treatment in which the student sees a teacher do something and then imitates that behavior.

These systems, which demonstrate an emphasis on skill development under simulated conditions, can clearly produce a gain in skills without the trainee having direct field experience. Although I am neutral on the question of performance-based teacher education, I feel that if we are going to make progress in the teacher education domain, we should place substantially greater emphasis on skill development in undergraduate training. Our present programs tend to be based too much on the verbalization of knowledge and not enough on the expression of knowledge in operable skills. A modest alteration in the programs along these lines would surely lead to a great deal of improvement in teacher education.

The question of field experience is more complicated. We have a tendency to think that field experience is always beneficial. This morning, one person in our

audience suggested that the practitioner ought to be doing all the training of teachers. There is some merit to this, but it is not entirely clear exactly what outcomes we would get unless we did it with a great deal of care. For instance, there is not much evidence to support the widely held assumption that teaching experience leads to increased skills. The analysis of the evidence we picked up in the 60's shows that the curve for skill acquisition in the teaching of reading and arithmetic rises very rapidly during student teaching, when the trainee is under supervision. It also rises rapidly the first two full years of experience, but here the gain is a function of the amount of supervision the teacher gets (Turner, 1965). By the amount of supervision, I mean the number of quarter-hours of contact supervisory personnel spend with new teachers. These personnel would be trained supervisors in reading or mathematics who know their areas thoroughly and make their living in this job – but interestingly enough, typically not principals. In the studies I am referring to, supervision tends to be the chief variable in explaining skill acquisition. If the teachers are not supervised, they apparently are not able to increase their skills very much.

Most teachers acquire the skills they are going to utilize by the fifth year and from then on, they probably practice what is most comfortable. There is an increment only if there is some kind of intervention. For example, one of our faculty members has developed an in-service intervention called "mastery learning" (Okey, 1972). This instructor provides the teachers with training booklets that are designed so teachers can complete each section at home in about one-half hour each (between TV programs). Okey has found that teachers do not respond favorably to these materials if the sections are longer, if only for the reason that the sections have to be worked in between other duties. The sections require the teachers to form objectives for their instruction, make diagnoses, prescribe further instruction as a consequence of diagnosis, collect their materials, engage in actual instruction, do remediation, and evaluate. In other words, the instructor is teaching each of these skills step-by-step to teachers. The inservice teacher then goes into the classroom and practices each of these steps or skills with a group of pupils.

Okey has used several kinds of evaluation designs with this mastery learning system. In one, he has had teachers divide their students into two groups, one of which is taught by the newly acquired teaching procedure, while the other group is taught by the teacher's usual procedures. The two groups are then compared. In another design, the teacher uses the mastery system in teaching one unit of instruction, but not in the next unit he or she teaches. For example, if the topic is fractions, the teacher will use the particular skill or skills to teach a unit on simple fractions, but not use the system to teach a more complex set of fractions. The outcomes of the two units are then compared. As a consequence of this intervention system, even with teachers who have been out in the field for fifteen years, student performance increases.

A field-based program, to be effective, must include resources for enough supervisors to have very close control over what happens during the field experiences of trainees. The program should not simply throw the trainees into the field

with a number of teachers and expect the trainees to develop a solid set of teaching skills.

One of the concerns of many people about performance-based teacher education is the amount of emphasis the instructor should place on pedagogical skill development and field experience. A variety of program options, each with different kinds of emphases or levels of emphasis, is one way to meet this concern. At Indiana University we have instituted alternative teacher education programs, each one different from the other. We have a multicultural program with a heavy field component in which, for example, trainees are working on Indiana reservations, teaching the children and learning about Indian culture. In another more traditional program, the trainees receive heavily knowledge-laden training after which they do student teaching. A third program emphasizes Protocol film and simulation material, and after training, conducts some performance testing. Judith Henderson's program at Michigan State also has a heavy element of simulation training related to skill acquisition. There is no one way to judge which of the three emphases is best. Performance-based programs should reflect a good deal of diversity in order to optimize specific skill, knowledge and attitude outcomes among trainees who vary in aptitude and in personal characteristics.

Certification

In certification, "program approval" is one option. It is the traditional option. The chief difficulty with this option is that one cannot be certain how much skill a particular individual actually has when he or she comes out of college. Can the person perform well in an actual classroom? The trainee will not have had very much classroom experience – perhaps only eight to twelve weeks of part-time work during student teaching. We could hardly expect anyone to acquire very high levels of skills during such a short period. But with PBTE program approval, which New York State already has, programs will probably be able to increase the amount of quality control over programs and thus speak more directly to questions of individual performance. Nonetheless, at the moment there is a poor research and development basis for such programs. People who specify that certain skills should be acquired are going beyond current knowledge since they cannot demonstrate that those skills make any difference in teaching. Initially at least, they are just hoping that they will. The development-evaluation paradigm which I mentioned earlier (see line D in the figure) will, however, permit teacher educators to test their hopes against actual evaluation data.

We should be very cautious about making promises about the delivery of even a "Program Approval" option in PBTE. It will probably take over ten years to find out whether PBTE is (or was) worth the investment. Only then will we have a chance to see a large number of trained teachers in different types of performance-based programs functioning in the field.

Individual approval is the other certification option available. A certification board would in some manner test an individual at a particular time on a performance basis and decide whether or not that person demonstrates adequate perfor-

mance. It is, however, very difficult for me to imagine a testing center comprehensive enough to give a fair test to an individual performer over a short period of time. Recently, when I was in Syracuse, some people there were discussing the possibility of a teacher counseling and improvement center. Here, after different kinds of measures have been used to assess and diagnose what a teacher can do, the teacher would try to improve those skills needing attention. This approach is different from attempting to certify someone by testing them at a skill center since it diagnoses deficiencies and attempts to correct them rather than to evaluate them for purposes of certification. In all, state approval of performance based teacher education programs, is probably the least noxious option in this domain.

Work Success Criteria — Process Skills

Applying process criteria involves observing the acts of the teacher and judging whether those acts are performed at an adequate level. The problem with process criteria, even if we find correlations between teacher performance and pupil performance, is the appropriate or optimum level of teacher skill performance. In a performance-based program, we must ask what degree of skill or level of behavior somebody has to engage in to be considered skillful. But how are we going to judge this from correlational studies? It is a question of the standards we apply. We ultimately cannot avoid making judgments against these standards. The problem is that research studies do not tell us the appropriate standards or levels for specific skills. As a matter of fact, there is reasonable evidence that some behaviors are best practiced in the middle range — not too much or too little.

It might be thought that if we greatly emphasized a particular behavior or skill, we would get more achievement gain from students because the relationship is seen monotonic and linear, always increasing. However, a behavior might have bad effects if over-emphasized. A study by Robert Soar suggests that teachers who allow too much physical freedom in the classroom get unfavorable results, whereas teachers who do not allow enough physical freedom in the classroom also get unfavorable results (Soar, 1973). The question is, how much is enough? I cannot answer that question. I am not even sure I know how to answer it. New York State, however, is certainly going to be faced with questions of this kind within the next year or so in one form or another. What are the standards to be applied to teacher performance?

Product Criteria

As for product criteria, McDonald has already presented some factors he has examined in pupil achievement. I would like to add to that some points on professional judgment and student reports. We obtained some rather surprising results from a study we conducted at Indiana in first-year French courses. For two years, we looked at all of the teachers in beginning French — eighteen one year and twenty-four the next year. In these courses, we measured student residual gain — the amount of learning as evidenced on the final exam corrected for the initial or entry level of learning. We also looked at the performance on the mid-term

test and the third-quarter test. We then had the students rate the teachers at the end of the course. We found that teachers who were rated high had less student achievement (Turner and Thompson, 1974). In this context, this finding suggests that a program which has as criteria both pupil achievement data and pupil report data about teachers will run into problems because the two criteria are negatively related. We also found the largest factor in the rating is how well-liked and how warm the teacher was. In general, the warmer a beginning French teacher is, the worse his students do.

These results probably depend very much on the distribution of student ability in learning French. We obtained stronger relationships in the first year when we had all the students rate their teachers. We lost students the second year, and the relationships were a little weaker, but still significant. In the second semester of French, we had lost about fifty percent of the students because it seems that the students were not disposed to take the next semester, at least immediately. They may have learned a good deal the first semester, but they apparently did not like it. In any event, the correlations for second semester were still weaker than those for first semester. A drop-off rate in a course may be inconsequential in public school, but it is not in colleges which have tuition-paying students. A French Department determines its number of faculty positions in considerable part by how many student credit hours it has. Clearly it has a problem when it loses fifty percent of its students between the first and the second semesters. The department is thus faced with a bad set of choices – shall it lose enrollment because of unpopularity in its teaching or shall it go for increased student gain? These are difficult choices to make.

I want to add here that in public schools it is indeed serious what students report about their teachers or their schools. Bad reports reflect on bond issues, tax levies, and other matters that duly influence the financial conditions of a school system. The relationship, however, is not as direct as in colleges where tuition income is directly involved.

We have some other results that are tied to the relationship between process criteria and product criteria. McDonald's study is going to produce some of these relationships; however, there have been a good many studies in the field, and I doubt that his study will be conclusive. One of these variables is "indirectness," which in the Flanders system means the frequency with which the teacher accepts pupil feeling, approves pupil behavior or gives approving responses, and responds to or expands on student questions. In general, "indirectness" is a positive response to students.

Another variable concerns the amount of criticism and the amount of self-justification that a teacher goes through. "Indirectness," or the first three approval categories, is placed over the criticism and self-justification categories to give the i/d ratio. Using this ratio, researchers try to predict which teachers obtain the better learning from students. The results of these studies, as reviewed by Dunkin and Biddle (1974), indicate that in nine out of eleven studies, pupils liked the indirect teacher better than the direct teacher, while in 33 studies relating i/d ratio to

student achievement, fifteen studies showed a positive relationship, fifteen showed no relationship, and three indicated a complicated kind of relationship. Should instructors train teachers to be indirect? Maybe it helps and maybe it doesn't. Insofar as praise and approval are concerned, one study of those teachers who gave more approval showed better pupil attitudes and four did not show any relationship. With respect to student achievement and the amount of approval, three studies produced positive relationships and eleven showed no relationships at all. In studies of the relationship between acceptance of pupil ideas and student attitudes, three studies resulted in positive relationships and three showed no relationships. With respect to the relationship between the acceptance of student ideas and pupil achievement, one study was positive, and five studies showed no relationship. With respect to teacher criticism of students, a recent study by Brophy and Evertson indicated that middle class students learn more under critical teachers, while lower class students learn less.

These studies are bothersome from the viewpoint of PBTE because they deal with well-researched variables, and yet, pupil outcomes cannot be consistently predicted from them. We really do not know what kind of teacher behavior consistently relates to pupil achievement outcomes. Should we certify a person who is indirect and not certify the one who is direct? There is insufficient evidence to support advocating a particular set of skills in this domain.

Data on classroom management rests largely on the work of Jacob Kounin (1970), but unfortunately there has been little replication of his work. In classroom management, Kounin identified a variable called "withitness," which means the extent to which a teacher can keep several things under control in the classroom at the same time. Teachers working with small groups would be able to monitor, say, the reading group with their right eye and all the pupils doing seat work with their left eye, and with the eye in the back of the head still be able to see what is going on in the rest of the room. In the work done by Kounin, "withitness" was correlated with another variable called "overlap," a behavior in which the teacher is able to do two or more things at once. The degree to which a teacher possesses this skill is a key factor in whether the pupils in the class are on a task or show deviant or off-task behavior. For overlap to be a relevant variable, however, we first have to assume that the particular task assigned to the pupils or the task they themselves have undertaken is important behavior. If we agree on that, then the amount of overlap the teacher shows will be an important correlate of the on-task behavior of the pupils.

Using a test built by Schumm (1971) on "overlap" some fairly interesting results with student teachers were obtained. The following is an example of an item in this test. A slide showing children in a classroom situation with their hands up is briefly presented. At the same time, the school principal reads an announcement over the loudspeaker. Afterwards, the teacher is instructed to write down both how many children had been pictured with their hands up and what the principal had said. Interestingly, the results of this test predicted that the elementary school student teachers who had more overlap had fewer discipline problems. The results also indicated that teachers who liked to work with multiple small groups rather than a

single large group scored higher in overlap behavior. However, this study showed a peculiar correlation to the studyhall behavior of teachers – the less overlap they had, the better off they were in supervising students in studyhall. This happens, perhaps, because studyhalls are a unidimensional activity. Teachers in studyhall do not have very many things to manage simultaneously.

Other important characteristics of classroom management include "smoothness" or the flow of teacher behavior, group learning in which the teacher makes sure the group knows what he or she wants them to do, and accountability, or making kids responsible for the consequences of their own behavior. Valence or challenge arousal – or the ability to stimulate enthusiasm and keep pupils involved is another important behavioral variable. This behavior, which is called stimulus variation in the Stanford system, requires the teacher to use a variety of materials and many different activities to keep the pupils' attention on what they should learn or what the teacher is saying or doing.

Turning now to behavior modification, there are several highly specialized studies which support the notion of targeting reinforcement to get a particular behavior. Let us suppose I have targeted a class of behaviors that I want to change in someone. For example, I want to get a pupil I'll call Jane to smile more often, so I would address positive reinforcement to that behavior. Everytime Jane smiles, I would say, "That's good, Jane" or perhaps just smile back. Pretty soon, Jane smiles more. But if Jane raises her left eyebrow, I do not say, "That's good," I don't say anything. Targeted praise is probably a more effective positive reinforcement than general praise.

The best systems for determining a teacher's positive and negative responses are the rapid-marking type systems, as in Flanders' CIA, where the observer rates a response once every three seconds, and the time-sampling scales in which the observer looks at teacher behavior and then checks the appropriate categories once every five to ten minutes.

Incidentally, the question of the reliability of observation is best handled by inter-observer agreement. For example, if you have three observers in the room checking the frequency of smiling, and they all produce the same rating, then we have what is technically known as "objectivity" in that behavior, as validated by inter-rater agreement. If only one person observes the behavior, we can assume that this observer's perception is subjective and cannot be verified. As the McDonald study indicates, inter-observer reliability is most easily done by multiple observers viewing a videotape of a teacher.

The number of different teacher behaviors and skills which can be listed in an observation schedule or category system is, of course, very large. I have noted above only a few, predominantly those with some research history behind them. The broad question facing the creators of performance-based teacher education programs is to find a way to select the skills to be included in the program when only a few have at this time been shown to have any claim to validity. In constructing the *General Catalog of Teaching Skills* we tried to confront this question squarely. Our resolution of it is described below.

A Differentiated Approach to Teaching Skills

The question of identifying teacher skills functional in a pluralistic society, and also observations of what goes on in schools in a wide variety of contexts, led us not to specify a single undifferentiated list of skills in constructing the *General Catalog*. Rather, we first divided the catalogue according to subject matters, adding separate sections for "Socialization and Classroom Management" and for "Early Childhood," which is an age level and not really a subject matter. As a second step, we used Bruce Joyce's idea (Joyce & Weil, 1972) that within subject matters teachers use different approaches to or different models of instruction. Here we tried to reflect schools of thought within different subject matters. For example, some segments of society believe that corporal punishment is moral, and others, that it is immoral. As researchers, we are confronted with these two opposing value systems. In the catalog, we tried to represent accurately a pluralistic system in a fairly objective way.

In the chapter dealing with "Socialization and Classroom Management," we came up with four schools of thought, one of which we called the authoritative approach. For teachers who take this approach, we included skills in administering corporal punishment, with the idea that if pupils are to receive corporal punishment, the teacher ought to give it in a skillful way. Many other skills related to "keeping discipline" are, of course, included.

The second approach, social-developmental, which has a long history in this country, and is identified primarily with Progressive Education, stands in substantial contrast to the authoritative approach. The cognitive development approach, the third approach employed, is derived from Piaget and Kohlberg, and deals with the open-classroom style of management and socialization. The fourth school of thought is the non-directive, or represents the Carl Rogers–Virginia Axline, approach. The skills to be acquired under this approach are very, very different from those identified with the authoritative approach.

Within a subject matter, a teacher deals with several functions including goal-setting, arranging the environment and selecting materials. Many studies have focused on another function – interaction or face-to-face instruction. However, in a system Lawrence Stolurow invented several years ago, there was substantial emphasis on what he called the pre-tutorial or planning behaviors – materials selection, programming behaviors, and so forth – that go into teaching, and less emphasis on the interactive aspect. Thus we identified these pre-tutorial functions as involving important skills which are frequently overlooked. Diagnosis and Evaluation are also functions of teachers. Most teachers assign grades; they also try to figure out why certain of their pupils are not doing as well as they should be doing. After considering these various functions of teachers, we decided to divide all approaches within subject matters according to the type of skills needed to perform each function.

Next we divided the skills according to the types of environments in which they could be observed. One type of environment is the college classroom and learning laboratory. The other context is micro-teaching or the actual classroom. Finally we

listed indicators for each skill under those contexts. For example, a skill under "Non-Directive Teaching" is "establishes and maintains rapport with students." Let us take a look at how we might test that in the college classroom or learning lab by looking at certain indicators. We could give the trainee a transcript of teacher-pupil interaction and a form on which he or she could code teacher talk as rapport-building versus threatening or rejecting. Or, we could show the trainee a film or videotape of teacher-pupil interaction, after which he or she would code the same behaviors using Flanders' categories 1, 2, 3, 6 and 7. In microteaching or in an actual classroom, our list of indicators for "Establishing and Maintaining Rapport" would include: greeting individual students in a relaxed way, complimenting or praising students, making humorous, non-threatening remarks, soliciting student opinion or comments, listening to student observations or opinions and encouraging expansion, making observations or comments which indicate acceptability of student expression; and, avoiding the following behaviors: constantly passing judgment on student behaviors, dress, friends or family, derogating or criticizing students, frowning, constantly carping, and being physically withdrawn or aloof. Some of these behaviors are positive indicators of the skill that we are building and some are negative indicators. Trainees can code these actions, either from video-tape, to show that they understand the elements of the skill we are building, or from a direct observation, to determine whether the teacher has the rapport-building skills.

If we decide to measure a skill like "Establishes Rapport," we must clearly identify the behaviors that make up or indicate that skill. How would I know, for example, that a teacher is enthusiastic, or that he or she can select appropriate materials? Unless we define and agree on the same behaviors in a skill category, we can never establish communication. When I search for the behaviors that constitute a particular skill, I am not attempting to control these behaviors, I am only trying to identify their components. In a pluralistic system, the programs, organizers or the instructors make the value judgments. The researcher's job is to develop the materials that permit the instructors to express their own values.

The *General Catalog* contains about fifty skills per section, but these behaviors are not independent of each other. When we take functions across subject matters, we get a repetition of skills, but in a different form. A question in teaching math is whether the teacher can evaluate certain kinds of mathematics behavior in children. We repeat the question in a different form for English teaching. In other words, these are similar behaviors with different content. A teacher might be appraised as skillful in evaluating math performance and incompetent in evaluating performance in English. Yet, there is a commonality in these teacher skills because both involve some kind of evaluative behavior. In essence, the situation appears to be this: the functions performed by teachers such as goal setting, arranging the learning environment, interacting with pupils, or diagnosing and evaluating are general or cut across subjects and age levels. Teaching skills, however, are specific since they involve operating precisely within a subject matter while instructing a particular level of student. Thus, all teachers perform the same general functions, but the skills they exercise may be viewed as specialized according to what is taught and who is being taught.

Under the concepts used to organize the *General Catalog*, the question of the general definition of teacher *skillfulness* can be resolved by noting that the less skillful teacher has control over only a few skills under a single approach to teaching a specific subject or a specific age level. A highly skillful teacher, on the other hand, is one who has mastered all of the skills necessary to perform each of several approaches within the subject matter taught. For example, a highly skilled social studies teacher is not merely skilled in the "assignment-recitation" approach to teaching social studies, but also has full control over the social inquiry approach, the social participation approach and other approaches to instruction useful as teaching contexts vary.

Summary

Teacher education has a definite structure which incorporates selection, training, certification, and placement, and finally, work success. At each point in this structure several program options or choices may be identified. The research base for selecting one option over another is modest, although some pertinent information is available from research in teacher education and teacher behavior during the past 20 years. The urgent question of which teaching skills and antecedent knowledges might best be incorporated into teacher education programs cannot be answered very well on the basis of the available research information. For this reason, the alternate approach of the *General Catalog of Teaching Skills*, which is based on approaches to, or models of, teaching, is proposed as one workable way to resolve this question.

BIBLIOGRAPHY

Baker, Lambert W., "An Analysis of Some Assumed Predictors of Success in Teaching," Doctoral Thesis, United States International University, 1970.

Balzer, David M., "The Influence of Selected Personal and Situational Characteristics on Shifts in Preference for Teacher Education," Doctoral Thesis, University of Minnesota, 1966.

Brophy, Jerry E., and Evertson, Carolyn, "Texas Teacher Effectiveness ..oject: Summary Report," AERA Paper, Chicago, April, 1974.

Ducharme, Robert J., "Selected Pre-service Factors Related to Success of the Beginning Teacher," Doctoral Thesis, Louisiana State University, 1970.

Dunkin, M. J. and Biddle, B. J., *The Study of Teaching*, New York: Holt, Rinehart and Winston, 1974.

Hiller, Jack H., "Verbal Response Indicators of Conceptual Vagueness," *American Educational Research Journal*, Vol. 8, No. 1, 151-161, January, 1971.

Joyce, Bruce and Marsha Weil, *Models of Teaching*, Englewood Cliffs: Prentice-Hall, 1972.

Kounin, Jacob, *Discipline and Group Management in Classrooms*, New York: Holt, Rinehart and Winston, 1970.

Okey, James and Ciesla, Jerome, *Teaching for Mastery*, National Center for the Development of Training Materials in Teacher Education, Indiana University, 1972.

Perry, James Z., "The Influence of Selected Factors on the Choice of Teaching as a Career," Doctoral Thesis, University of Oklahoma, 1972.

Schumm, Robert W., "Performance on Multiple Attention Measures as a Predictor of the Classroom Management Proficiency Displayed by Student Teachers," Doctoral Thesis Indiana University, 1971.

Soar, Robert S., "An Integrative Approach to Classroom Learning," NIMH Project 5-R11 M11 01096, University of South Carolina: Temple University, 1960.

Trojcak, Doris Ann, "Five Stages of Instruction for Sequencing Science Activities according to Gagne's Learning Model," Doctoral Thesis, Indiana University, 1969.

Turner, Richard, *The Acquisition of Teaching Skills in Elementary School Settings*, Bulletin of the School of Education, January, 1965.

Turner, Richard, Editor, *A General Catalog of Teaching Skills*, Final Report for Grant No. OEG-0-71-1070, Indiana University, December, 1973.

Turner, Richard, White, K. W., Quinn, E. H., and Smith, N., *Skill in Teaching: Three Studies*, Bulletin of the School of Education, January, 1964.

Turner, Richard, "Some Predictors of Problems of Beginning Teachers," *Elementary School Journal*, February, 1967.

Turner, Richard L. and Thompson, Robert P., "Relationships between College Student Ratings of Instructors and Residual Learning," AERA paper, Chicago, April, 1974.

Weeden, Shirley U., "Screening for Teaching," *Improving College and University Teaching*, 124, Spring, 1972.

Yoder, Samuel L., "An Investigation of the Relationships between Student Teaching Problems and Selected Student Teacher Characteristics," Doctoral Thesis, Indiana University, 1967.